DOCUMENT RESUME

ED 187 724 TM 800 219

AUTHOR Froman, Robin D.; Owen, Steven V.
TITLE Influence of Different Types of Student Ratings Feedback Upon Later Instructional Behavior.
PUB DATE Apr 80
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980). Best copy available.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Check Lists; *College Faculty; *Feedback; Higher Education; Information Utilization; Pretests Posttests; Rating Scales; Research Needs; *Student Evaluation of Teacher Performance; Teacher Effectiveness; *Teacher Improvement; Teaching Styles
IDENTIFIERS University of Connecticut

ABSTRACT

The short-term effect of feedback of student ratings of college teachers was investigated, but no significant differences were found. Ratings of 84 teachers, in three experimental groups and a control group, were obtained in the third week of the fall semester and again during the final week of the same semester. Feedback in the form of a computer printout was provided by the sixth week of the semester. One group received feedback in the form of scores on ten high-inference rating scales describing broad aspects of teaching effectiveness. Another group received feedback regarding their position on items from a 68-item behavioral check list. The third experimental group received both kinds of feedback, and the control group received no feedback. No significant differences were found among the four groups with respect to either pretest ratings or demographic information. Multivariate analysis of covariance failed to indicate any significant differences among the four groups with respect to posttest ratings. (The evaluation form is appended, as well as sample printouts). (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187724

BEST COPY AVAILABLE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Influence of Different Types of Student Ratings Feedback Upon Later Instructional Behavior

Robin D. Froman

Steven V. Owen

University of Connecticut, CT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robin D. Froman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM800 219

A paper presented at the annual meeting of the American Educational Research Association, 1980, Boston, MA.

INTRODUCTION

The evaluation of teaching has received increased attention from educational researchers and theorists alike over the past twenty years. Various means of teacher evaluation have been explored. Three categories of teacher evaluation are currently recognized and used in educational settings. The three types include presage, product and process evaluation (Mitzel, 1960). Presage evaluation involves the inspection of characteristics of teachers or potential teachers in order to attempt to predict teaching effectiveness. Presage characteristics may include such things as college grades, number of years of teaching experience, and personal appearance. Product evaluation involves the inspection of student outcomes which are thought to be the result of effective instruction. Such variables might include student grades in a course, student achievement on standardized tests, or student achievement scores over a period of time. Process evaluation involves the use of observation of the actual teaching process in order to identify teacher or student behaviors thought to indicate effective instruction. One type of process evaluation is the use of student rating scales of teaching; this is the focus of the research reported here.

Student ratings at the postsecondary level are traditionally used to serve a number of purposes. The purposes include the provision of: a) information regarding teacher effectiveness to be used in making promotion and tenure decisions; b) information for students to use in making decisions about course and instructor selection; and c) a source of remedial feedback to be used by teachers to improve instruction (Costin, Greenborough, and Menges, 1971).

The use of student ratings as one of the three most common criteria for making promotion and tenure decisions is well documented (Bejar, 1975). While publications and service to the community (the two other criteria used in making such decisions) are deemed important contributions by any faculty, they do not reflect the primary responsibility of teaching professionals.

That primary responsibility is the instruction of students. Some have asserted that student ratings of instruction are currently the predominant means by which teaching effectiveness is evaluated in higher education (Trent and Cohen, 1973).

The second purpose, as a source of information for students, is a more controversial issue. Because of questions related to the objectivity of student ratings (Owen, 1976), and because of occasional opposition of teachers' unions to the use and publication of student rating results (Remmers, 1963; Owen, 1976) the availability of publications of student ratings of teachers has been limited.

The final use of student ratings--as a source of feedback to be used by teachers--is the primary concern of the present research. The proponents of student rating instruments have often assumed that feedback of ratings to teachers necessarily helps to make instruction more effective. Kulik and McKeachie (1975) point out, however, that there is no convincing evidence that information about students ratings improves teaching. Assumptions about changes in instructional effectiveness have been based on limited and equivocal evidence.

Types of Student Rating Scales

Another intuitive assumption seems to have been made about different types of feedback being equally effective in producing changes in teacher behavior. Two broad forms of feedback are available. One form of feedback is general in nature and based upon student responses to high inference rating scales. High inference rating scales include broad items which require a great deal of interpretation by the rater between observing behaviors and recording a rating. For example, students' ratings of their teacher's "attitude toward students" would require considerable inference from myriad teaching behaviors to a single rating response. The second form of feedback is more specific and is based upon low inference rating scales. Low inference scales include items which require little or no interpretation on the part of the student between observing a behavior and rating that behavior. An item such as "instructor meets class regularly" is an example of a low inference item. In this item there is little abstraction from an assortment of behaviors to a single rating.

It is also an assumption that feedback from low inference items can be more helpful in altering teacher behavior (Lolli, 1977). Presumably, this is because of a specificity and relative precision of low inference items. If one is told that he "presents information at too fast a pace," the message is clear. By contrast, if one is rated low on the high inference item, "stimulates interest," how shall he improve? The inference gap is thus seen to operate in two ways: between the students' perception and their ratings, and between the instructor's feedback and his interpretation of what it means. Imprecision and error interfere with the usefulness of high inference items when students try to reduce a set of teacher behaviors into a single rating and again when instructors try to relate ratings on broad items to specific instructional behaviors. Through the use of low inference scales, it may be possible to eliminate the steps of changing specific observations to high inference ratings and then changing those ratings back to specific recommendations for instructional change. In his presidential address to the American Psychological Association McKeachie (1976) suggested that student ratings must provide instructors with useful information in order to facilitate improvements in teaching. Surely information is most useful when it is both clear and objective. Unfortunately, the relative efficacy of low inference feedback has not been suitably tested and remains an hypothesis.

Overview of Related Literature

A plethora of studies dealing with student ratings of instruction have focused on reliability and validity issues related to the scales. A review by Costin et al. (1971) presents an early representative sample of these studies. Internal consistency and stability estimates reported in the review ranged from about .70 to .94 with a majority of the estimates clustering in the .80s. Validity estimates reported by Costin et al. varied more widely than did the reliability estimates, but were summarized as indicating that "students can provide reliable and valid information on the quality of courses and instruction" (p. 530). Other more recent studies are typically supportive of validity and reliability issues (see Linn, Centra, and Tucker, 1975; Centra, 1976; Lolli, 1977).

Unlike reliability and validity issues, the questions regarding the effectiveness of feedback from rating scales have been given limited attention. Until the 1960s the effectiveness of feedback to teachers as a remedial tool was taken on faith alone. Citing no empirical evidence, McKeachie (1969)

suggested that most teachers would agree that feedback of student ratings of instruction improves teaching. Rosenshine and Furst (1973) asserted that one of the primary uses of rating data from teacher evaluation instruments is to train prospective teachers and to improve their teaching.

Some objective data pertaining to the influence of feedback of ratings have been generated. Gage, Runkel, and Chatterjee (1963) reported that when feedback of ratings given by sixth grade students was provided to teachers, subsequent teacher ratings improved for a portion of the teachers. An experimental study involving high school and technical institute teachers and students showed that teachers receiving feedback declined less (!) in subsequent ratings than did teachers receiving no feedback (Tuckman and Oliver, 1968).

Miller (1971) was unable to replicate the positive effects of ratings feedback at the college level. After supplying feedback to graduate teaching assistants and comparing subsequent ratings of teaching with a no feedback control group, no differences were found in later ratings. Centra (1973) extended the research on the effects of feedback at the college level to include additional variables. Although Centra reported no significant differences between the student ratings of feedback and control groups of college faculty after a half semester interval, differences did emerge after longer periods of time. This research indicated that the amount of time elapsed between feedback and later ratings may contribute to changes.

In a similar study, Braunstein, Klein, and Pachla (1973) found improved performance by college teachers, as reflected by student ratings, following feedback of student evaluations. Braunstein et al. delivered low inference feedback to teachers, and reported positive changes in later ratings for teachers who overestimated their initial student ratings. However, because of the small sample size (N=19) and the use of weak analytic techniques, the results must be regarded with caution.

In an examination of the effects of student feedback coupled with a discussion of how comments could be implemented in the classroom, Friedlander (1978) reported students were most likely to note changes if adequate discussion followed ratings. When looking at the outcomes of ratings and varying degrees of discussions in 85 classes, students seemed to attribute

positive changes in instruction to the discussion of student ratings. This finding implies that written comments, when supplemented by discussion and clarification, led to meaningful change. Unfortunately, a direct test of instructional improvement was not undertaken in the study. Students were only asked to report whether they thought discussion of ratings was sufficient and if the discussions led to change.

In their review of the studies dealing with the effectiveness of students' evaluative feedback, Rotem and Glasman (1979) summarize some of the recommendations which are frequently encountered. In particular, they found numerous calls for provision of feedback on specific instructional behaviors and the use of low rather than high inference rating scales. These recommendations come not only from the studies reviewed, but from Rotem and Glasman's own synthesis of the results of the past twenty years of research. They suggest that the equivocal results of the feedback studies may be explained by a number of variables, one of them being the specificity of the feedback delivered.

A summary of the ratings literature tends to support the reliability and validity of student ratings of instruction. The research regarding the effects of feedback of ratings is not nearly so extensive or definitive. Replications of studies reporting improvements in teaching following feedback are non-existent. Because of the many differences in types of feedback provided it is impossible to make a generalizable statement about whether feedback leads to positive instructional changes. If summed across the various types of feedback delivered, the evidence of instructional change following feedback at the university level is minimal (Rotem and Glasman, 1979).

Several specific problems can be identified as limiting the literature related to the question of the effectiveness of feedback. In the majority of studies reviewed, univariate techniques of analysis were employed. Often, raw change scores were used as dependent measures (Tuckman and Oliver, 1968; Pambookian, 1974). Restricting analyses of rating data to univariate techniques interferes with determining complex, multivariate relationships between type of feedback and instructional change. Other problems apparent in some of the research on feedback are the use of small samples and, again, the lack of replication. Unless a large or particularly representative sample is used, or unless a study can be replicated, confidence in the results

must remain limited. A final problem limiting the research on feedback of ratings of teaching is the relative absence of experimentally controlled investigations of the comparative influences of clearly defined high versus low inference feedback. Although such studies have been recommended (Rotem and Glasman, 1979), few can be found. Existing research is usually quasi-experimental at best; correlational, descriptive strategies seem to dominate the literature.

Objectives

Beyond assumptions and educated guesses, the differential effectiveness of high versus low inference feedback upon subsequent instruction is difficult to document. Even those few research studies which investigate the effects of either high or low inference feedback in isolation are commonly limited by the use of univariate techniques. No research study currently exists using multivariate techniques and a large, representative sample to simultaneously compare the effects of low inference, high inference, a combination of low and high inference, and no feedback given to instructors at the post secondary level. If differential effects could be determined as the result of type of feedback, the feedback most effective in improving teaching techniques could be identified. Once identified, such feedback could be delivered to instructors to remediate instructional problems and to improve teaching in general.

The purpose of this research was to investigate whether or not feedback influences later instructional practices of college faculty as evidenced by subsequent student ratings. The research was directed toward identifying the differential effects of type of feedback supplied to instructors--high inference, low inference, a combination of high and low inference, or no feedback--as it contributed to later ratings of instruction. The feedback issue fits within the broader context of instructional improvement as inferred from improved student ratings.

The research attempted to meet the following objectives:

1. to determine if post ratings of instruction on high inference items are affected by type of feedback delivered (i.e., low inference feedback, high inference feedback, both types of feedback, no feedback); and,

2. to determine if post ratings of instruction on low inference items are affected by type of feedback delivered (i.e., low inference feedback, high inference feedback, both types of feedback, no feedback).

INSTRUMENTATION AND PROCEDURES

The instrument used to collect data was the University of Connecticut Survey of Courses and Teaching (see Appendix A). Reliability estimates for the scale were reported by Lolli and Owen (1976). The internal consistency estimates for the factors underlying the scale ranged from .55 to .86 with a mean estimate of .73. Content validity was supported by the review and opinions of a panel of expert judges. Construct validity was partially supported by the emergence of meaningful and interpretable factors underlying the scale (Lolli, 1977).

The instrument is divided into three major sections. The first brief section contains questions about student demographic data. The second section, items 1 through 11, is representative of typical high inference items used on teacher ratings instruments. Responses to these items were used as the source of data to provide the general, high inference feedback to instructors. The third section is a diagnostic checklist of 68 items used to provide the low inference, specific feedback to instructors.

Sample

Over one hundred instructors who taught during the Fall, 1977 semester at the University of Connecticut-Storrs agreed to cooperate with the research effort following a request to participate in the study. Selection of those requested to participate was accomplished by identifying every fifteenth instructor scheduled to teach a course as listed in the University of Connecticut Schedule of Courses for the Fall of 1977. Instructors were not informed as to the precise nature of the research, but were told that the study was part of continued efforts to demonstrate the validity and utility of the University of Connecticut Survey of Courses and Teaching.

Of the total group of volunteering instructors, 88 were identified as teaching classes appropriate for inclusion in the research study. Team

taught classes, field instruction courses, and classes with fewer than five students were not included in the sample. Due to administrative difficulties,[1] four of the original 88 instructors and their classes were dropped from the analytic sample. Thus, a total of 84 instructors and their classes were included in the final sample.

Research Design

During the third week of the Fall 1977 semester, the students in the classes taught by the participating instructors were requested to complete, anonymously and in class, the University of Connecticut Survey of Courses and Teaching for their respective instructors.

Class median ratings were calculated for each of the 11 high inference evaluative items on the scale (item 1 through 11; see Appendix A). Medians were calculated rather than means because of the tendency for ratings on the scale to be negatively skewed (Lolli, 1977). Class medians were used as the unit of analysis in all of the statistical procedures. Medians were also used to develop written feedback of ratings on the high inference, evaluative items. Frequency counts of the number of students checking each of the low inference, diagnostic items (included in the "Diagnostic Check List"; see Appendix A) were used to supply written feedback of ratings on the low inference, specific items.

The instructors in the sample were assigned randomly to one of the four experimental groups. The four groups were identified as follows:

Experimental Group 1 - received feedback of their students' ratings on the 11 high inference rating items on the scale;

Experimental Group 2 - received feedback of their students' ratings on the 68 low inference rating items on the scale;

[1]These problems included such things as cancelled classes or snowstorms occurring on the day scheduled for data collection.

Experimental Group 3 - received feedback of their students' ratings on the 68 low inference items and the 11 high inference rating items on the scale; and

Control Group - received no feedback of their students' ratings of their instruction.

Feedback was delivered to those instructors in Experimental Groups 1, 2, and 3 in written form by the sixth week in the semester (see Appendices B-E for samples of the feedback delivered to instructors).

During the final week of classes in the Fall 1977 semester, all students in the participating classes were again asked to complete the University of Connecticut Survey of Courses and Teaching during a class period.

Statistical Procedures

Preliminary analyses were conducted in order to determine the initial equality of the four randomly assigned groups. A multivariate analysis of variance (MANOVA) was conducted using the vector of the initial class ratings on the 11 high inference rating items as the dependent measure, and group membership (either Experimental Groups 1, 2, 3 or Control Group) as the independent measure. A second preliminary MANOVA was conducted to parallel the first, except that the dependent vector employed was based upon the 68 low inference items.

Additional analyses directed at identifying any initial differences between the four experimental groups included a series of one-way ANOVAs. These analyses were conducted using various demographic characteristics of the instructors as the dependent variables and group membership (one of the four experimental groups) as the levels of the independent variable. The demographic variables included sex of instructor, years of teaching experience, percent of time employed, age, and tenure status.

Two sets of analyses were conducted to determine the influence of feedback of student ratings. A series of one-way multivariate analyses of covariance (MANCOVAs) were employed to identify differences between the treatment groups on final ratings. Group membership was the independent variable

in each of these analyses. Median end-of-semester ratings on the eleven high inference evaluative items formed the vector of dependent measure in one MANCOVA. That MANCOVA was used to determine the answer to the first research question: does feedback of student ratings influence subsequent teacher behavior as evidenced by high inference ratings?

The second research question was addressed through ten MANCOVAs using the treatment groups as the independent variable and groups of the low inference items as the dependent vectors. Research reported by Owen, Froman and Lolli (1976) documents the relationship between the eleven high inference items and the 68 low inference items on the University of Connecticut Survey of Courses and Teaching. During validation of the scale median class ratings on each of the evaluative items were calculated for each instructor in the sample. A median split was then imposed on the data to identify those instructors receiving class ratings higher or lower than the total group median on each of the evaluative items. Group membership, based upon the median split for each of the high inference items, was the independent variable in a series of one-way ANOVAs. The dependent variables in these ANOVAs were selected low inference items. Item selection was based upon the judgments of three content experts. The experts identified low inference diagnostic items which were conceptually related to each of the high inference items. Groups of low inference items were identified for ten of the eleven high inference items. One of the evaluative items, "overall teaching ability," was not included in these analyses. The content experts suspected that all of the low inference items contributed to a rating on that single high inference item. Thus, a limited number of low inference items could not realistically be determined for that evaluative item.

For each of the ten high inference items considered by the experts a series of ANOVAs were conducted. They were used to determine if there were differences in the low inference ratings awarded to the two groups of instructors. The results of the ANOVAs, adapted from the tables presented by Owen et al., are presented in Appendix F. The results of these analyses have been used in the present research to group the low inference items. Ratings on the ten empirically identified groups of the behaviorally oriented, low inference items formed the dependent vectors in the ten MANCOVAs used to answer research question two: does feedback influence subsequent ratings on low inference items? In all the MANCOVAs ratings collected during the third week of classes served as the covariate.

RESULTS

The analyses conducted to identify any initial differences between the experimental groups yielded nonsignificant results. The first preliminary analysis, a MANOVA comparing initial ratings among the four comparison groups showed the mean ratings vectors to be equivalent (F=.85, df=33,213). The second set of preliminary analyses, the ANOVAs conducted on the demographic variables, yielded similarly nonsignificant results. In those analyses, the instructors' age, sex, percent of time employed, number of months per year employed, number of years since hiring, tenure status and years tenured were the dependent variables. The respective F values in those seven analyses were 1.09, 1.84, 0.38, 0.76, 0.41, 0.26, and 0.61. In all the ANOVAs the df equaled 3,81. Based upon the lack of statistically significant differences in these variables before treatment, it was assumed that a multivariate covariance design could be used in analyzing post-treatment differences. The MANCOVA could be used without posing a threat to the validity of the interpretation by artifically controlling for variables which might have contributed to important differences. In this design, the covariate option in the multivariate analysis was used to account for any differences (even those which were not statistically significant) which might accumulate over the use of a series of variables and lead erroneously to posttreatment differences.

The MANCOVAs examining student ratings following treatment yielded nonsignificant results. In each case, the ratings vectors of the four treatment groups were statistically equivalent on both low inference and high inference end of semester ratings. Table 1 summarizes these outcomes.

CONCLUSIONS AND DISCUSSIONS

Based upon the results of the multivariate analyses, it appears that feedback alone, regardless of its form, has no short term influence upon teaching behaviors as perceived and reported by students. This finding suggests that the assumed differences in the outcomes of low versus high inference information do not occur. In addition, the results imply that one should not expect to observe substantial differences in teacher ratings over a single semester following the delivery of feedback of ratings.

These findings, while not statistically significant, do have significant implications for research in the area of student ratings. They provide important information pertaining to the influence of two types of feedback, low and high inference, when provided with no additional counseling and when observed over a short period of time. The results of this study confirm the results reported by Miller (1971) and Centra (1973), and contradict those reported by Braunstein et al. (1973). Because of the representativeness of size of the sample, and the use of multivariate analyses, many of the problems in research design which limited the generalizability of previous studies have been eliminated.

Two directions for research in student ratings must now be explored further. Longer range studies need to be conducted to determine longer range effects of ratings feedback. Centra (1973) has also suggested this type of research. Unfortunately, longitudinal studies are costly and difficult to implement. Such a study would have realistic applications, however, because of the nature of teaching. Existing teaching behaviors are practiced and repeated so frequently that they may be difficult to change. Any significant changes in behaviors may require time for practice and shaping of the new behaviors.

There is reason to believe that students' perceptions of teacher behaviors are fairly sturdy and resistent to change (Bausell and Magoon, 1972). The stability of students' percentions of teaching implies that two groups of students would be used to provide a baseline of ratings for the purposes of later comparisons and for feedback to teachers. The second group of students, taking the same course one semester or more later and following the provision of feedback of ratings to teachers, could provide teacher ratings to be used to determine if change had occurred.

The second area of research to be explored pertains to the impact of counseling about student ratings of teaching. Results reported by Jacoby (1977) indicate that individual consultation to discuss implications of ratings for subsequent teaching did lead to differences in later student ratings. Unfortunately, his results were severely limited by the sample size (N=4 in the treatment group) and the study needs to be replicated and extended. In addition, the cost effectiveness of providing individual consultation to

instructors needs to be evaluated in a representative, large sample in order to determine the feasibility of providing such a service to faculty in a college or university setting. Studies addressing these issues have been recommended by others (Rotem and Glasman, 1979) but have yet to be conducted.

In summary, the sample and design of the present study has served to shed some light on the contradictory findings previously reported in the area of student ratings. Feedback of student ratings, when delivered without consultation, appears to have negligible immediate impact upon teaching behaviors as perceived by students. The type of feedback provided, either low inference, high inference or a combination of the two, does not have a differential influence on subsequent ratings.

Table 1

Summary of Eleven MANCOVAs on Mean Vectors of Post-Treatment Ratings

Dependent Vector[1]	f	df
11 Evaluative Items	1.05	66,177
Diagnostics: Evaluative #2	1.04	102,141
Diagnostics: Evaluative #3	0.90	72,171
Diagnostics: Evaluative #4	0.69	18,212
Diagnostics: Evaluative #5	0.89	42,199
Diagnostics: Evaluative #6	1.20	18,212
Diagnostics: Evaluative #7	1.32	114,129
Diagnostics: Evaluative #8	1.21	12,204
Diagnostics: Evaluative #9	1.03	12,204
Diagnostics: Evaluative #10[2]	0.30	6,158
Diagnostics: Evaluative #11	1.06	60,182

[1]Appendix F gives the matching of Diagnostic items with Evaluative Items.

[2]Actually, this analysis was an ANCOVA, bacause there was only a single diagnostic item associated with Evaluative #10.

Bibliography

Bausell, R. and Magoon, J. Expected grade in course, grade point average, and student ratings of a course and instructor. Educational and Psychological Measurement, 1972, 32, 1013-1023.

Bejar, I.I. A survey of selected administrative practices supporting student evaluation of instruction programs. Research in Higher Education, 1975, 3, 77-86.

Braunstein, D. N., Klein, G.A., and Pachla, M. Feedback expectancy and shifts in student ratings of college faculty. Journal of Applied Psychology, 1973, 58, (2), 254-258.

Centra, J.A. Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 1973, 65 (3), 395-401.

Centra, J.A. The influence of different directions on student ratings of instruction. Journal of Educational Measurement, 1976, 13 (4), 277-282.

Centra, J.A. Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 1977, 14 (1), 17-24.

Costin, F., Greenough, W.T., and Menges, R.J. Student ratings of college teaching: reliability, validity, and usefulness. Review of Educational Research, 1971, 41 (5), 511-535.

Friedlander, J. Student perceptions on the effectiveness of midterm feedback to modify college instruction. Journal of Educational Research, 1978, 71 (3), 140-143.

Gage, N.L., Runkel, P.J., and Chatterjee, B.B. Changing teacher behavior through feedback from pupils: an application of equilibrium theory. In W.W. Charters and N.L. Gage (Eds.) Readings in the Social Psychology of Education. Boston: Allyn and Bacon, 1963.

Jacoby, K.E. Behavioral prescriptions for faculty based on student evaluations of teaching. American Journal of Pharmaceutical Education, 1977, 41, 8-13.

Kulik, A. and McKeachie, W.J. The evaluation of teachers in higher education. Review of Research in Education, edited by F.N. Kerlinger, p. 210-240. Itasca, Ill: F.E. Peacock, 1975.

Linn, R.L., Centra, J.A., and Tucker, L.R. Between, within, and total group factor analysis of student ratings of instruction. Multivariate Behavioral Research, 1975, 10, 227-288.

Lolli, A. and Owen, S.V. Student ratings: What is the frame of reference: A paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1976.

Lolli, A. A university-wide multivariate validation of a student rating scale for teacher evaluation. Unpulished dissertation, University of Connecticut, 1977.

McKeachie, W.J. Student ratings of faculty. AAUP Bulletin, 1969, 55, 439-444.

McKeachie, W.J. Psychology in America's bicentennial. American Psychologist, 1976, 12, 819-833.

Miller, M.T. Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 1971, 62 (3), 235-239.

Mitzel, H.E. Teacher effectiveness. In Harris, E.W. (Ed.) Encyclopedia of Educational Research, 3rd ed. N.Y.: MacMillan, 1960, p. 1481-1486.

Owen, S.V. The validity of student ratings: a critique. A paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1976.

Owen, S.V., Froman, R.D. and Lolli, A. Technical Report on the Development and Validation of the University of Connecticut Survey of Courses and Teaching. Storrs, CT: Bureau of Educational Research and Service, 1976.

Pambookian, H.W. Initial level of student evaluation of instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology, 1974, 66 (1), 52-56.

Remmers, H.H. Rating Methods in Research on teaching. In N.L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Rosenshine, B. and Furst, N. The use of direct observation to study teaching. In R.M.W. Travers (Ed.) Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973.

Rotem, A. and Glasman, N.S. On the effectiveness of students' evaluative feedback to university instructors. Review of Educational Research, 1979, 49 (3), 497-511.

Trent, J.W. and Cohen, A.M. Research on teaching in higher education. In R.M.W. Travers (Ed.) Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973.

Tuckman, B.W. and Oliver, W.F. Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 1968, 59 (4), 297-301.

Department _____ Course # _____ Section _____ Branch _____ Instructor _____

FOR OFFICE USE ONLY

DIRECTIONS:
PLEASE USE A NO. 2 PENCIL TO FILL IN APPROPRIATE BLOCKS. DO NOT MAKE STRAY MARKS. DO NOT SIGN YOUR NAME. ERASE CLEANLY IF YOU CHANGE AN ANSWER. RESULTS WILL NOT BE SEEN BY THE INSTRUCTOR BEFORE THE COURSE IS COMPLETED.

Semester Standing	Is Course in Your Major	Expected Grade in This Course	Cumulative Average (QPR)	How Often Do You Attend Class
1 = 1 - 2 2 = 3 - 4 3 = 5 - 6 4 = 7 - 9 5 = 9 OR MORE 6 = GRADUATE 1 2 3 4 5 6	YES NO	A B C D F PASS AUDIT	1 = LESS THAN 10 2 = 10 - 14 3 = 15 - 19 4 = 20 - 24 5 = 25 - 29 6 = 30 - 34 7 = 35 OR MORE 1 2 3 4 5 6 7	1 = 90 - 100% 2 = 75 - 89% 3 = 50 - 74% 4 = 25 - 49% 5 = 0 - 24% 1 2 3 4 5

Please omit those of the following items which are not appropriate:

In my opinion this instructor is excellent, good, average, less than average, poor with regard to:

	POOR	LESS THAN AVER.	AVER.	GOOD	EXCEL.
1. Overall teaching ability					
2. Presenting material in a clear and effective manner					
3. Demonstrating overall organization					
4. Making purpose and objectives of course clear					
5. Fulfilling course objectives					
6. Making work assignments and student responsibilities clear					
7. Stimulating interest					
8. Grading fairly and impartially					
9. Using examination items which stress important aspects of course					
10. Being accessible to students both in and out of class					
11. Showing an interest in and a concern for students					

DIAGNOSTIC CHECK LIST: The following is for the purpose of providing specific feedback to your instructor. Please mark ONLY those traits in the performance of your instructor which you found to be PARTICULARLY NOTABLE.

THE INSTRUCTOR

- is consistently late in starting or ending class
- is consistently unprepared
- has illegible handwriting
- has a distracting smoking habit
- seems unfriendly toward students
- is friendly toward students
- is fair to everyone
- criticizes and embarrasses students unfairly
- encourages student questions or comments
- is patronizing, talks down to students
- maintains attitude of "I am always right"
- enforces pointless rules
- seems preoccupied with personal problems
- is too dogmatic or opinionated to present materials fairly
- displays favoritism toward certain students
- displays prejudice on the basis of race, sex, religion, ethnicity
- speaks clearly and effectively
- speaks in a monotone
- is difficult to hear or understand (mumbles, accent, inaudible)
- has distracting speech habits
- has distracting mannerisms (nervousness, pacing, twitch, etc.)
- engages in too much idle chit-chat
- uses class time effectively
- just reads lectures
- presents interesting lectures; stimulating
- stimulates curiosity
- displays enthusiasm
- presents dull, boring lectures
- does not cover materials in depth, too superficial
- summarizes material effectively
- relates course to real life situation
- shows sense of humor
- presents too much detail or trivia
- sequence of materials is logical
- stresses important points
- does not stress important points
- presents material at too fast a pace
- presents material at too slow a pace
- goes over the same thing too often
- uses examples, illustrations well
- does not use enough illustrations, examples
- uses special materials or labs that were useful
- uses discussions effectively
- spends too little time on class discussions
- responds to questions in a vague and confusing manner
- answers questions completely
- relates readings and lectures clearly
- assigns reading materials and then ignores them
- readings and assignments are meaningful
- does not explain what is expected on exams or assignments
- does not provide enough feedback or comments on exams or papers
- provides helpful feedback
- is slow in returning graded work
- stresses rote memorization or trivia rather than understanding on exams
- applies grading standards inconsistently
- does not provide enough exams or other materials for evaluation
- requires purchase of costly materials which are not worth the price
- assigns too much "busy work"
- makes unreasonably long and burdensome assignments
- increases appreciation of subject matter

THIS COURSE

- is better than expected
- too much smoking takes place
- cheating is widespread
- enrollments are too large for effective instruction
- content is not sufficiently challenging
- content is over my head, too difficult
- content does not match expectations, catalog description
- too much responsibility is delegated to graduate assistant

NCS/NDF - 579:54321

APPENDIX A: University of Connecticut Survey of Courses and Teaching

Appendix B

ID # 1210

Dear Professor _____:

Below is a partial summary of your ratings from the first administration of the revised JCONN Rating Scale for Instruction. At this time the only ratings available from your class are those for the Evaluative items. Because of problems with scoring the new instrument additional data from the scale are not summarized at this time. We hope to have the summary of all of your ratings available and returned to you by January, 1978.

The items included in this summary are the eleven Evaluative items at the top of the scale. For each item the median rating is printed. Medians have been used because they are usually more representative of the students' impressions.

If you have comments or questions regarding the summary, or if you wish further clarification, please call Robin Froman, 486-2080.

We will return for the second round of evaluations in your class on _____ at _____.

Thank you for your help and cooperation with this project.

EVALUATION ITEMS (1=POOR/2=LESS THAN AVG/3=AVG/4=GOOD/5=EXCELLENT)	MEDIANS	# RESPONSES
Overall teaching ability	3.056	101
Presenting material in a clear and effective manner	3.798	101
Demonstrating overall organization	4.034	101
Making purpose and objectives clear	3.947	99
Fulfilling course objectives	3.958	92
Making work assignments and student responsibilities clear	3.854	98
Stimulating interest	3.375	100
Grading fairly and impartially	3.921	41
Using exam items which stress important aspects of course	3.611	30
Being accessible to students both in and out of class	4.058	84
Showing an interest in and a concern for students	4.167	91

ID # 2460

Dear Professor _____:

Below is a partial summary of your ratings from the first administration of the revised UCONN Rating Scale for Instruction. At this time the only ratings available from your class are those for the Diagnostic Check List items. Because of problems with scoring the new instrument additional data from the scale are not summarized at this time. We hope to have the summary of all of your ratings available and returned to you by January, 1978.

The items included in this summary are those from the Diagnostic Check List. Only those items which were checked by at least 25 percent of the students as being 'particularly notable' in your class are listed. Also included with these items is the actual percentage of students who happened to check the items.

If you have comments or questions regarding the summary, or if you wish further clarification, please call Robin Froman, 486-2080.

We will return for the second round of evaluations in your class on _____ at _____.

Thank you for your help and cooperation with this project.

Diagnostic Check List Items	% Checked
Uses class time effectively	49.35
Is friendly towards students	46.75
Is fair to everyone	33.77
Summarizes material effectively	45.45
Relates course to real life situation	36.36
Shows sense of humor	55.84
Sequence of materials is logical	35.06
Stresses important points	42.86
Uses examples, illustrations well	55.84
Enrollments are too large for effective instruction	28.57
Speaks clearly and effectively	64.94
Assigns reading materials and then ignores them	25.97

NUMBER OF STUDENTS RESPONDING = 77

ID # 3150

Appendix D

Dear Professor _____:

Below is a summary of your ratings from the first administration of the revised UCCVM Rating Scale for Instruction. For each of the eleven Evaluative items the median rating is printed. Medians have been used because they are usually more representative of the students' impressions.

You may also find that one or more of the Diagnostic items are printed below the Evaluative items. The reason these items appear is that at least 25 percent of the students in your class checked the items as being 'particularly notable' in your class. Also included with these items is the actual percentage of students who happened to check the items.

If you have comments or questions regarding the summary, or if you wish further clarification, please call Robin Froman, 486-2080.

We will return for the second round of evaluations in your class on _____ at _____.

Thank you for your help and cooperation with this project.

EVALUATION ITEMS (1=POOR/2=LESS THAN AVG/3=AVG/4=GOOD/5=EXCELLENT)	MEDIANS	# RESPONSES
Overall teaching ability	3.875	12
Presenting material in a clear and effective manner	3.875	12
Demonstrating overall organization	4.167	12
Making purpose and objectives clear	4.333	12
Fulfilling course objectives	4.100	10
Making work assignments and student responsibilities clear	4.000	12
Stimulating interest	3.667	11
Grading fairly and impartially	5.000	0
Using exam items which stress important aspects of course	5.000	1
Being accessible to students both in and out of class	4.500	9
Showing an interest in and a concern for students	4.300	12

Diagnostic Check List Items	% Checked
Readings and assignments are meaningful	50.00
Uses class time effectively	58.33
Displays enthusiasm	41.67
Is friendly towards students	83.33
Summarizes material effectively	50.00
Relates course to real life situation	66.67
Encourages student questions or comments	75.00
Shows sense of humor	58.33
Sequence of materials is logical	50.00
Stresses important points	66.67
Increases appreciation of subject matter	41.67
This course is better than expected	33.33
Uses discussions effectively	41.67
Speaks clearly and effectively	75.00
Relates readings and lectures clearly	41.67

NUMBER OF STUDENTS RESPONDING = 12

The University of Connecticut

STORRS, CONNECTICUT 06268

SCHOOL OF EDUCATION

Dear Professor

Due to problems encountered while trying to score and summarize your ratings from the first administration of the UCONN Rating Scale for Instruction we are unable, at this time, to provide feedback of your ratings. We apologize for this delay and hope to be able to provide a full summary of your ratings from both administrations of the scale by late December or early January, 1978.

Someone will return to your classroom on _____ at _____ to administer the rating scale for a second time.

Thank you very much for your help and cooperation in this project and your tolerance of our delay in de-bugging our machinery. If you have any questions or comments please call me at 486-2080.

Thank you again for your help.

Sincerely,

Robin Froman

RF/amf

Appendix F

Low Inference Items Related to

High Inference Evaluative Item #2 *

(Presenting material in a clear and effective manner.)

---

1. presents dull, boring lectures
2. responds to questions in a vague and confusing manner
3. does not stress important points
4. is difficult to hear or understand (mumbles, accent, inaudible)
5. goes over the same thing too often
6. does not cover materials in depth, too superficial
7. speaks in a monotone
8. does not use enough illustrations, examples
9. presents too much detail or trivia
10. is patronizing, talks down to students
11. seems preoccupied with personal problems
12. is consistently unprepared
13. presents material at too fast a pace
14. is too dogmatic or opinionated to present materials fairly
15. has illegible handwriting
16. has distracting speech habits
17. has distracting mannerisms (nervousness, pacing, twitch, etc.)

---

* each low inference item is statistically related to the high inference item with which it is grouped at the $p < .05$ level.

Low Inference Items Related to

High Inference Evaluative Item # 3*

(Demonstrating overall organization)

---

1. does not explain what is expected on exams or assignments
2. does not stress important points
3. presents too much detail or trivia
4. assigns reading materials and then ignores them
5. spends too little time on class discussions
6. does not cover materials in depth, too superficial
7. is consistently unprepared
8. is slow in returning graded work
9. does not provide enough feedback or comments on exams or papers
10. seems preoccupied with personal problems
11. goes over the same thing too often
12. presents material at too fast a pace

---

* each low inference item is statistically related to the high inference item with which it is grouped at the $p < .05$ level.

Low Inference Items Related to

High Inference Evaluative Item #4 *

(Making purpose and objectives of course clear)

---

1. does not stress important points
2. does not explain what is expected on exams or assignments
3. stresses rote memorization or trivia
4. rather than understanding on exams

Low Inference Items Related to

High Inference Evaluative Item #5 *

(Fulfilling course objectives)

---

1. does not stress important points
2. does not cover materials in depth, too superficial
3. presents too much detail or trivia
4. engages in too much idle chit-chat
5. presents material at too slow a pace
6. is consistently unprepared
7. presents material at too fast a pace

---

*each low inference item is statistically related to the high inference item with which it is grouped at the $p < .05$ level.

Low Inference Items Related to

High Inference Evaluative Item #6 *

(Making work assignments and student responsibilities clear)

---

1 does not explain what is expected on exams or assignments

2. is consistently unprepared

3. assigns reading materials and then ignores them

Low Inference Items Related to

High Inference Evaluative Item #8 *

(Grading fairly and impartially)

---

1. applies grading standards inconsistently

2. displays favoritism toward certain students

Low Inference Items Related to

High Inference Evaluative Item #9 *

(Using examination items which stress important aspects of course)

---

1. stresses rote memorization or trivia rather than understanding on exams

2. does not provide enough exams or other materials for evaluation

---

* each low inference item is statistically related to the high inference item with which it is grouped at the $p < .05$ level.

Low Inference Items Related to

High Inference Evaluative Item # 7 *

(Stimulating interest)

1. presents dull, boring lectures
2. presents too much detail or trivia
3. goes over the same thing too often
4. responds to questions in a vague and confusing manner
5. does not stress important points
6. is difficult to hear or understand (mumbles, accent, inaudible)
7. speaks in a monotone
8. enforces pointless rules
9. just reads lectures
10. does not cover materials in depth, too superficial
11. is too dogmatic or opinionated to present materials fairly
12. maintains attitude of "I am always right"
13. has distracting speech habits
14. spends too little time on class discussions
15. stresses rote memorization or trivia rather than understanding on exams
16. is patronizing, talks down to students
17. does not use enough illustrations, examples
18. displays prejudice on the basis of race, sex, religion, ethnicity
19. assigns too much "busy work"
20. displays favoritism toward certain students
21. seems unfriendly toward students
22. makes unreasonably long and burdensome assignments

* each low inference item is statistically related to the high inference item with which it is grouped at the $p < .05$ level.

.14.

Low Inference Items Related to

High Inference Evaluative Item # 10 *

(Being accessible to students both in and out of class)

1. seems unfriendly toward students

Low Inference Items Related to

High Inference Evaluative Item # 11*

(Showing an interest in and a concern for students)

1. maintains attitude of "I am always right"
2. does not explain what is expected on exams or assignments
3. enforces pointless rules
4. is patronizing, talks down to students
5. seems unfriendly toward students
6. does not provide enough feedback or comments on exams or papers
7. criticizes and embarrasses students unfairly
8. makes unreasonably long and burdensome assignments
9. requires purchase of costly materials which are not worth the price
10. assigns too much "busy work"

*each low inference item is statistically related to the high inference item with which it is grouped at the $p < .05$ level.